Finding and using routine clinical datasets for observational research and quality improvement

Lucy M. McDonnell,
NIHR In Practice Research Fellow,
School of Population Health and Environmental Sciences,
King's College London
3rd Floor, Addison House
Guy's Campus
LONDON
SE1 1UL
Lucy.mcdonnell@kcl.ac.uk

Brendan Delaney,
Chair in Medical Informatics and Decision Making
Faculty of Medicine, Department of Surgery & Cancer
Imperial College, London
Queen Elizabeth the Queen Mother Wing (QEQM)
St Mary's Campus
Brendan.delaney@imperial.ac.uk

Prof. F.M. Sullivan,
Prof. of Primary Care Medicine
Head of Division Population & Behavioural Science
University of St. Andrews

Gordon F. Cheesbrough Research Chair, North York General Hospital.
Professor, Department of Family & Community Medicine and
Dalla Lana School of Public Health, University of Toronto.
Adjunct Scientist Institute for Clinical Evaluative Sciences(ICES)
fms20@st-andrews.ac.uk

Introduction

Primary care in the U.K. generates an extraordinary amount of data. There are more than 300 million consultations annually, creating unrivalled opportunities for research.(1) The volume of patients that consult primary care practitioners daily, the variety of clinical conditions, the diversity of populations and the transfer from hand written records to comprehensive electronic medical systems has heralded a new era in primary care research. Furthermore, the linkage of primary and secondary care data systems creates opportunities

for prospective and retrospective studies and epidemiological insights into population health.(2)

Increasing accessibility of rich data has changed the landscape of research in the community. As well as large datasets based around electronic medical records, primary care researchers also have access to alternative sources of data, which are often free, and record linkage amongst datasets in safe havens can enhance the value of records further.(3) The aim of this article is to highlight datasets which are available to primary care researchers and to give examples of how they have been used in primary care research. A new resource detailing primary care/community based datasets is now available. This resource has been developed by the Farr Institute, an organisation which aims to build capacity in health informatics research (www.farrinsitute.org). The resource is a catalogue of U.K based datasets with metadata (data which provides information about other data) which may be useful to novice and experienced researchers in primary care.

Datasets in catalogue

The catalogue has been divided into the following categories: electronic medical record data, quality of primary care services, prescribing data, audit, health surveys, special datasets, cohort studies, administrative dataset and screening datasets. Available metadata include type of data, context and method of extraction, coverage, geography, duration, volume, granularity (level of detail), coding, consent and access (including websites and contact details), and were reviewed by dataset custodians. A brief overview of some categories and examples of how the datasets have been used in research follows below.

Electronic Medical Record data

This section includes large national datasets such as Clinical Practice Research Datalink (CPRD) and QResearch, regional datasets such as the Secure Anonymised Information Linkage Databank (SAIL) based in Wales, and local databases such as Lambeth DataNet which all use electronic medical records based on computer systems. Some datasets can be linked with secondary care data to carry out cross-sectional or cohort studies. A recent example is a cohort study which used data from the QResearch database linked to the

national cancer registry, to develop and validate risk prediction equations to estimate survival in patient with colorectal cancer.(4)

Quality of primary care services

UK Quality and Outcome Framework (QOF) data is routinely collected by GP surgeries. Martin et al used QOF data to look at recording of physical health targets of those with major mental illness compared with those with chronic kidney disease across the UK. Their findings suggested inequality in access to certain aspects of health care for patients with major mental illness.(5)

Prescribing data

Regional prescribing data is available across the UK and is often used in research studies, looking at cost-effectiveness of interventions or prescribing patterns. Ashworth and his colleagues were able to show that reduced antibiotic prescribing in general practice was associated with decreased patient satisfaction, by linking national patient survey data and prescribing data for England.(6)

Audit

Although the majority of audits are based in secondary care, some audits are based in the community. The National Audit of Cancer Diagnosis in Primary care, for example, has been used to study the variation of promptness in presentation of patients subsequently diagnosed with cancer.(7)

Health Surveys

Health surveys for England, Scotland, Wales and Northern Ireland are carried out on annual basis and provide rich data on the health of the nation. Results from the Scottish Health survey were used to show the relationship between dental health and cardiovascular disease mortality.(8) Each country also has a national patient cancer experience survey which have been used to look at regional variations in cancer patient experience.(9)

Special datasets

An example of a special dataset is the Aberdeen Maternity and Neonatal Databank, which collects data from primary and secondary care. Lee et al used this dataset to look at maternal obesity during pregnancy and its association with major cardiovascular events in later life.(10) Linking the data with the national register of deaths and Scottish Morbidity Record, they were able to determine that maternal obesity is associated with increased risk of premature death and cardiovascular disease.

Cohort studies

A number of cohort studies exist at national and regional level which collect patient data in the community. The largest cohort study is the UK Biobank which holds data on 500, 000 participants and has been used for a large variety of research studies. Recently, Flint and Cummins used Biobank data to confirm the association between active commuting and healthier bodyweight and composition, supporting the case for promoting active travel to prevent obesity in later life. (11)

Screening datasets

National Health Service Screening datasets for each UK country are available for breast, cervical and bowel cancer screening. Massat et al used screening data to look at variation in cervical and breast cancer screening coverage in England, determining the effect of deprivation, ethnicity and urbanisation on screening uptake. (12)

Accessing and contributing to the catalogue

The catalogue of datasets has recently been developed and is currently available as a PDF document but will hopefully be available on an interactive digital platform in the future. It can be accessed via the following web link: http://www.farrinstitute.org/wp-content/uploads/2017/10/Datasets-that-may-be-of-interest-to-Primary-Care-Researchers-in-the-UK-May-2016.pdf. Further contributions from dataset custodians who would like their dataset to be included in the catalogue are encouraged to contact the Farr Institute primary care working group http://www.farrinstitute.org/research-education/research/primary-care.

1. Hippisley-Cox Julia, Yana V. Trends in consultation rates in General Practice 1995/1996 to 2008/2009: Analysis of the QResearch® database
2. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA. 2014 Jun 25;311(24):2479-80.
3. Lea NC, Nicholls J, Dobbs C, Sethi N, Cunningham J, Ainsworth J, et al. Data Safe Havens and Trust: Toward a Common Understanding of Trusted Research Platforms for Governing Secure and Ethical Health Research. JMIR Med Inform. 2016 Jun 21;4(2):e22.
4. Hippisley-Cox J, Coupland C. Development and validation of risk prediction equations to estimate survival in patients with colorectal cancer: cohort study. BMJ. 2017 Jun 15;357:j2497.
5. Martin JL, Lowrie R, McConnachie A, McLean G, Mair F, Mercer S, et al. Physical health indicators in major mental illness: data from the Quality and Outcome Framework in the UK. Lancet. 2015 Feb 26;385 Suppl 1:S61.
6. Ashworth M, White P, Jongsma H, Schofield P, Armstrong D. Antibiotic prescribing and patient satisfaction in primary care in England: cross-sectional analysis of national patient survey data and prescribing data. Br J Gen Pract. 2016 Jan;66(642):e40-6.
7. Keeble S, Abel GA, Saunders CL, McPhail S, Walter FM, Neal RD, et al. Variation in promptness of presentation among 10,297 patients subsequently diagnosed with one of 18 cancers: evidence from a National Audit of Cancer Diagnosis in Primary Care. Int J Cancer. 2014 Sep 01;135(5):1220-8.
8. Watt RG, Tsakos G, de Oliveira C, Hamer M. Tooth loss and cardiovascular disease mortality risk--results from the Scottish Health Survey. PLoS One. 2012;7(2):e30797.
9. Saunders CL, Abel GA, Lyratzopoulos G. What explains worse patient experience in London? Evidence from secondary analysis of the Cancer Patient Experience Survey. BMJ Open. 2014 Jan 03;4(1):e004039.
10. Lee KK, Raja EA, Lee AJ, Bhattacharya S, Norman JE, Reynolds RM. Maternal Obesity During Pregnancy Associates With Premature Mortality and Major Cardiovascular Events in Later Life. Hypertension. 2015 Nov;66(5):938-44.
11. Flint E, Cummins S. Active commuting and obesity in mid-life: cross-sectional, observational evidence from UK Biobank. Lancet Diabetes Endocrinol. 2016 May;4(5):420-35.
12. Massat NJ, Douglas E, Waller J, Wardle J, Duffy SW. Variation in cervical and breast cancer screening coverage in England: a cross-sectional analysis to characterise districts with atypical behaviour. BMJ Open. 2015 Jul 24;5(7):e007735.